众包在证据合成中的实践应用研究

——以 Cochrane Crowd 公民科学项目中的众包应用为例

李 晓[1,2]，曲建升[1,2,3*]，寇蕾蕾[4]

（1. 中国科学院西北生态环境资源研究院，兰州 730000；2. 中国科学院大学 经济与管理学院图书情报与档案管理系，北京 100049；3. 中国科学院成都文献情报中心，成都 610041；4. 兰州大学 青藏高原人文环境研究中心，兰州 730000）

摘 要：[目的 / 意义]证据生成的及时性对于循证决策至关重要，而目前证据合成的效率通常不能满足决策者的需求。众包被认为是一种可以提高证据合成生产效率的潜在方法。本研究以 Cochrane Crowd 公民科学项目中的众包应用为例，总结众包在证据合成中的实践应用。[方法 / 过程]采用文献调研、网络调查、案例分析等方法，从众包者、志愿者、众包任务、Cochrane Crowd 平台、质量评估 5 个维度分析了众包在 Cochrane Crowd 公民科学项目中的应用机制。[结果 / 结论]通过设置明确目标、激励措施、清晰任务，提供全面培训和适当的质量控制机制，可以应用众包为证据合成输出高质量结果。为未来针对不同领域证据合成中应用众包以及在证据合成的不同阶段使用众包的进一步研究提供参考。

关键词：证据合成；众包；Cochrane Crowd；循证研究

中图分类号：G254　　文献标识码：A　　文章编号：1002-1248（2023）02-0095-12

引用本文：李晓，曲建升，寇蕾蕾. 众包在证据合成中的实践应用研究——以 Cochrane Crowd 公民科学项目中的众包应用为例[J]. 农业图书情报学报，2022，35(2)：95-104.

1 引 言

循证决策（Evidence-Informed Decision-Making）是弥补研究与政策之间差距的一种手段[1]，而证据合成（Evidence Synthesis）已经成为很多领域循证决策的重要工具。对现有证据的准确、简洁和无偏见的合成是学术界为政策制定者或任何决策者提供的最有价值的贡献之一[2]。然而伴随着人类研究的广度和深度大约每 9 年翻一番[3]，进行证据合成是资源密集型的，尤其是针对复杂问题的科学知识进行合成，因为解决方案往往分布于数以万计的个体研究中。有研究表明单一的证据合成需要花费研究团队 18 个月到 3 年不等的时间[4,5]，这阻碍了政策的需求驱动周期，特别是对于需要答案

收稿日期：2023-01-21

基金项目：中国科学院战略性先导科技专项（A 类）“丝路环境科技态势监测分析与知识集成服务”（XDA2010030802）

作者简介：李晓（1983- ），博士研究生，中国科学院西北生态环境资源研究院，研究方向为循证方法、知识发现与知识组织。寇蕾蕾（1991- ），博士，助理研究员，兰州大学青藏高原人文环境研究中心，研究方向为知识发现与知识组织、数字人文、青藏高原人文环境研究

*通信作者：曲建升（1973- ），博士，教授，研究员，博士生导师，中国科学院大学，研究方向为战略情报分析、情报咨询与知识挖掘。Email：jsqu@lzb.ac.cn

才能作出相关资源分配的决策。学术界在不断探索使用创新技术来提高证据合成生产效率，一些学者研究了通过机器学习、自然语言处理、文本挖掘等技术来提高效率[6]，一些学者研究了应用众包等公民科学方法来降低成本、缩短时间[7-13]，本研究主要关注众包在证据合成中的实践应用情况。Cochrane 是致力于健康领域证据生产的国际知名非营利组织，主要制作基于随机对照试验（RCT）的系统评价。虽然 Cochrane 主要为医学领域生产证据，但其开发的证据合成方法、指南、工具随后被其他领域借鉴和参考，因此 Cochrane 已经成为证据合成领域的先驱者和引领者。Cochrane[14] 较早在证据合成中使用了众包，并于 2016 年推出基于众包模式的 Cochrane Crowd 公民科学平台[15]。目前，对在证据合成中应用众包的研究和实践尚处于起步阶段，对 Cochrane Crowd 项目中众包的应用机制进行全方位分析，将为在证据合成中使用众包提供一定的参考和启发，从而促进证据合成生产效率的提高，为循证决策提供及时而有力的科学信息。

证据合成涉及到对同一研究问题的多个研究的信息组合，遵循透明性、客观性、可重复性原则，在特定时间点提取特定问题的已知信息以形成总结和理解[16]。证据合成中最早使用且最常使用的方法是系统评价，该方法最开始被应用于健康领域（已成为该领域的黄金标准，基于随机对照试验的系统评价被视为最高级别证据），随后被其他学科广泛采用。为响应政策制定者以及其他利益相关者对信息类型和时效性的需求，更多的证据合成方法被开发和应用，包括范围评价、系统图、快速评价、实时系统评价、系统评价再评价等。其他证据合成方法主要建立在系统评价方法基础之上，因此这些方法的流程大致相同，它们之间的区别主要体现在每个步骤的具体操作上。证据合成的流程可以归纳为 7 个步骤：①提出研究问题并形成协议；②文献搜索；③资格筛选；④数据提取；⑤质量评估；⑥合成数据；⑦汇总证据并形成报告。

众包能够利用大量在线人群在更大范围内收集、分析或处理数据，在提升生产能力、效率及节约成本等方面具有优势，因而被认为是一种可以提高证据合成生产效率的潜在方法。理论上，人群可以为证据合成的各个阶段作出贡献。在搜索阶段，由于通常要检索多个书目数据库以及灰色文献来源，可以由每名参与者负责 1 个数据库，从而提高效率；在资格筛选阶段，参与者可以基于研究纳入 / 排除标准，对引文或全文作出纳入或排除的决策；在数据提取阶段，参与者可以帮助提取关于样本、干预措施、结果等特征信息；在质量评估阶段，参与者可以帮助评估偏倚风险、研究质量等；在合成阶段，参与者可以将数据输入到元分析软件中、进行元分析；在汇总证据、报告结果阶段，参与者可以编写报告和更新结论。但是在实践中，人们往往聚焦于将众包应用于证据合成中最耗费资源的阶段，正如 HADDAWAY 等所言，“最耗时的程序应该被视为方法和技术发展的重要领域，以提高效率”[4]。

通过调研已发表的关于证据合成中资源使用情况的文献，表明在证据合成步骤中，资格筛选、数据提取和质量评估最为耗费资源[4,17]。另外这些步骤所需的时间与检索结果数量、需要识别的全文数量、纳入研究数量有着密切的关系。国外学者已经对在证据合成中使用众包开展了一些可行性研究和实践应用，在这些研究与实践中，人群主要是参与了一个特定阶段，而不是完整的过程。大部分研究对在资格筛选阶段应用众包展开研究[7,9-11,13,18-21]，个别研究对在数据提取阶段和质量评估阶段应用众包展开研究[8,12]。

在证据合成众包的实践应用方面，规模最大、实施情况良好的是由 Cochrane 组织发起的基于众包模式的 Cochrane Crowd 公民科学项目。自 2014 年以来，Cochrane 一直使用众包为 CENTRAL 数据库有效识别健康证据，并于 2016 年 5 月推出了 Cochrane Crowd 公民科学平台，面向全球招募志愿者帮助对健康决策所需的研究进行分类。迄今为止（2023 年 3 月 10 日），已有来自全球 180 个国家或地区的 29 004 名贡献者，对将近 800 万条记录进行了分类。该项目没有让志愿者执行系统评价的所有步骤，而是专注于研究识别——基于标题和摘要确定文献是否满足特定类型的研究设计（如随机对照试验）。该项目中，识别证据的方式从传统的基于同行评议的孤立模式转变为充分利

用众多人员的协同模式，代表了寻找和管理研究信息方式的重要转变，证实了众包在证据合成中可以发挥非常实际的作用[22]。本研究采用文献调研、网络调查、案例分析的方法，以众包在 Cochrane Crowd 项目中的应用机制为例，总结众包在证据合成中的实践应用，以期为中国各个学科领域在证据合成中应用众包、提高证据生成效率提供借鉴和参考。

2　众包在 Cochrane Crowd 中的应用机制

在科学领域，学者纷纷提到了众包的维度，分别有一个维度，如 NAKATSU 等基于任务特征将众包分类为：合同聘用、分布式问题解决（附加 / 集合协调）、新创意生成、协作（互惠协调）[23]；两个维度，如 PRPIĆ 等提出了众包的两个维度：内容（客观 / 主观）和贡献（聚合 / 过滤）[24]；3 个维度，如 ESTELLÉS-AROLAS 等确定了众包的 8 个特点，可分为三大类：人群（谁、做什么，以及得到什么回报），发起者（谁、从人群中得到什么回报），过程（过程类型、呼吁类型、使用的媒介）[25]；4 个维度，如 HOSSEINI 通过对计算机科学、商业和管理、医学、环境科学和社会学等领域众包相关文献的分析，发现构成了整个众包运作的四大维度或支柱：人群、众包者（发起者）、众包任务、众包平台[26]；ZHAO 等认为众包有 4 个基本维度：提供者（一般人群或特定群体）、所有权（公共 / 私人物品）、模式（集体、竞争或合作）、动机和激励（内在 / 外在）[27]；6 个维度，如 PEDERSEN 等提出了众包的 6 个要素：问题、过程、治理、人（问题所有者、个人或群体）、技术和结果[28]。本文在 HOSSEINI 等构建的众包运作的 4 个维度基础上，增加了一个维度“质量评估”，将从众包者（发起者）、志愿者（人群）、众包任务、Cochrane Crowd 平台（众包平台）、质量评估五个维度来分析众包在 Cochrane Crowd 中的应用机制。

2.1　众包者

众包者即众包发起者，也称为请求者，可以是个人、机构、组织或公司。在 Cochrane Crowd 项目中，众包者有两种类型：Cochrane 组织和 Cochrane 系统评价作者团队，前者为该项目的发起者，也是项目的主要服务对象，后者为该项目推出 Screen4Me 服务后出现的发起者。

（1）Cochrane 组织。Cochrane 是由研究人员、专业人员、患者、护理人员以及对健康研究感兴趣的其他人员共同组成的全球性非营利组织，旨在收集和分析健康领域现有最佳证据并生成系统评价，以帮助人们对健康和卫生保健作出明智决策，其工作被公认为是高质量、可靠信息的国际黄金标准。Cochrane 主要制作基于随机对照试验的系统评价，内容涉及健康服务和健康政策研究，Cochrane 评价由于严格性、客观性，在证据中享受较高的声誉。为了提高系统评价生成效率，Cochrane 构建了对照试验中心数据库（CENTRAL），该数据库专门收集随机和半随机对照试验报告（RCTs 和 qRCTs）。Cochrane 组织发起众包的目的是通过人群的集体努力，为 CENTRAL 数据库识别来自于 Embase.com、CINAHL（护理学数据库）、ClinicalTrials.gov（美国临床试验注册中心）和 WHO ICTRP（世卫组织国际临床试验注册平台）等数据库的随机和半随机对照试验研究（RCT 和 qRCT）。

（2）Cochrane 系统评价作者团队。该项目于 2019 年推出了 Screen4Me 服务，来自 15 个 Cochrane 评价小组的 60 多个系统评价团队已经使用了该服务[19]。该服务由 3 个组件构成：已知评估服务、RCT 分类器服务以及 Cochrane Crowd 众包服务。Cochrane 系统评价作者团队使用 Screen4Me 的流程为：将特定主题系统评价的待筛选记录上传至平台后，首先通过已知评估服务与平台中已由人群筛选过的记录进行匹配，然后再通过 RCT 分类器服务对其余记录进行 RCT 与非 RCT 的识别，最后对潜在的 RCT 记录通过众包服务由志愿者帮助进一步识别。一些试点研究表明，志愿者不仅可以识别潜在的 RCT，还可以根据纳入标准评估 RCT 及其他类型研究是否与评价主题相关[19-21]，同时在这些任务中，需要单独开发定制培训模块。

2.2 志愿者

Cochrane Crowd 对所有人开放，只需要使用姓名和电子邮箱地址进行注册和登录。Cochrane 作为证据合成的知名国际组织为 Cochrane Crowd 提供了良好的志愿者基础，而 Cochrane Crowd 平台与该组织的其他 IT 基础设施集成在一起，也为志愿者参与其活动提供了便利。调查显示该项目志愿者的参与动机主要有利他主义（如帮助 Cochrane）、技能提升（如学习相关知识）、实现自我价值等。

Cochrane Crowd 为志愿者设计了绿、棕、银、金、紫五色里程碑徽章，分别代表完成训练任务、100 个分类、500 个分类、1 000 个分类以及出类拔萃（至少 1 000 个分类且精度很高），完成任务时相应徽章将出现在任务界面，以此激励志愿者完成更多的分类任务。另外，平台为每位志愿者提供了“众包活动总结”，可以看到自己参加学习活动或任务的历史记录。

根据任务完成情况，参与者被分为 3 个级别：普通筛选者、专家筛选者和解析者。每名志愿者都从普通筛选者开始，当以非常高的准确度完成 1 000 个分类（其中不确定分类的比例非常低）时可以升级为专家筛选者。专家筛选者所做的分类被赋予更大的权重，因而由专家筛选者筛选的记录需要较少的决策便可获得分类结果。解析者是具有出色筛选绩效的志愿者，当其他筛选者对记录的分类发生分歧或记录被分类为不确定时由解析者对这些记录作出最终决策，通常需要获得全文才能决定。由于所需专业水平较高，解析者通常人数较少。

2.3 众包任务

为了提升人群参与度、鼓励人群参与，Cochrane Crowd 没有限制志愿者的加入资格，允许没有任何健康研究经验的个体参与。为了使志愿者能顺利执行众包任务，采取了通过多途径提高志愿者的认知水平同时降低任务执行难度的策略。具体表现在精心设计了以任务为中心的学习活动、任务访问方式、培训模块和反馈机制等，并对众包任务本身进行了科学的类型划分和设计。

2.3.1 学习活动

提供了关键概念、理解研究设计和介绍 CONSORT（报告临床试验的统一标准）等学习活动，旨在帮助志愿者学习相关知识。关键概念部分需要完成 7 个小模块，掌握公平试验的一些关键概念；研究设计模块介绍了健康领域研究人员主要使用的一些研究设计类型；CONSORT 模块旨在介绍 CONSORT 的标题和摘要，讲授研究人员开展随机对照试验时所报告的信息。

2.3.2 任务访问方式

由于志愿者具有不同经验水平，因此设计了 3 种任务访问方式：直接进入任务、新人途径和学生途径。直接进入任务适合具有领域知识和专业经验的志愿者，新人途径适合对健康研究和循证医学比较陌生的志愿者，学生途径适合正在学习健康领域相关知识并希望对循证医学有更多了解的志愿者。当选择直接进入任务时，界面中会呈现所有的学习活动和任务类型，志愿者可以任选其一。新人途径和学生途径分别由一系列学习活动和任务依据循序渐进原则按照一定的顺序组合在一起，志愿者只能依次参加，如表 1 所示。

2.3.3 培训模式和反馈机制

每类型任务都有简短的格式相同的培训模块支持，该模块由不同数量的交互式培训记录组成（如 RCT 和 CT 识别中有 20 条培训记录，ICTRP 识别任务中有 7 条培训记录），在志愿者对记录作出分类选择后都有及时的反馈及引导。这些记录反映了志愿者在实时任务中可能遇到的情况、使其了解每个特定任务的基本内

表 1 新人途径和学生途径

Table 1 Newcomers' pathway and students' pathway

途径类别	学习活动和任务的名称及顺序
新人途径	关键概念、CT 识别、理解研究设计、介绍 CONSORT、RCT 识别
学生途径	关键概念、CT 识别、RCT 识别、Screen4Me、理解研究设计、介绍 CONSORT

容并通过反馈机制指导志愿者应如何完成。对于随机试验识别任务，在完成培训记录后可以直接执行任务，对于 COVID 研究识别任务，培训部分由学习和评估两部分组成，完成培训记录后还需要进行评估，正确率达到 7 条记录（共 10 条）时才可以执行任务，否则需要重复学习直到评估合格。

2.3.4　任务类型

Cochrane Crowd 中的众包任务可以分为三大类：主流任务、扩展任务和试点任务，各大类任务下又包括具体的任务类型，一些任务只有在志愿者完成 100 条 RCT 识别记录才可用，任务详情如表 2 所示。Cochrane Crowd 按照“页面”组织任务，每页中显示一条记录（由一篇文献的标题和摘要组成）以及 3 个标签：RCT/qRCT、Reject（拒绝）或 Unsure（不确定），用户通过选择其中一个标签对记录进行分类，完成选择后才能继续下一条记录。

主流任务为根据文献的标题和摘要识别随机对照试验（RCT）。需要志愿者识别的记录来源主要有 4 个：Embase.com、CINAHL、ClinicalTrials.gov 和 WHO ICTRP，不同来源的记录格式有所区别，识别难度也不尽相同。主流任务依据记录格式又细分为 3 类任务：RCT 识别、CT 识别和 ICTRP 识别。RCT 识别任务在用户界面中显示每篇文献的标题和摘要或仅标题（图 1），CT 识别和 ICTRP 识别任务在界面中显示一组结构化字段、信息更加清晰（图 2），对初学者比较友好和容易，被正确识别的可能性更高。

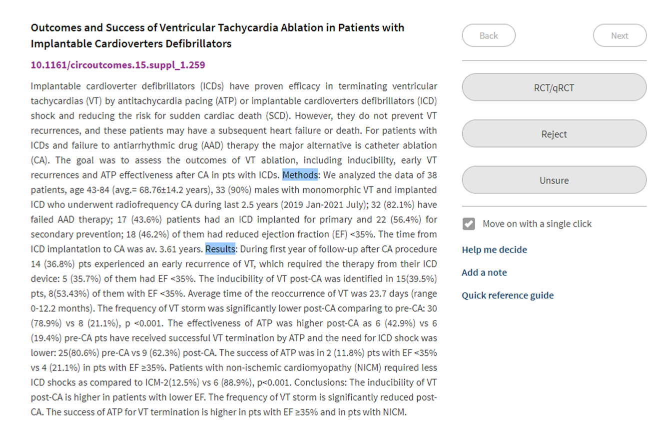


图 1　RCT 识别任务示例

Fig.1　RCT identification task

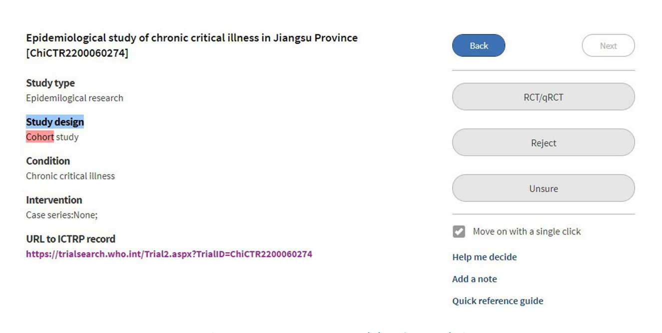


图 2　ICTRP 识别任务示例

Fig.2　ICTRP identification task

表 2　任务类型详细情况

Table 2　Details about task types

任务分类	任务名称	任务描述	任务难度	任务可用条件
主流任务	RCT 识别	帮助从 Embase.com、CINAH 等文献数据库获取的记录中找到随机和半随机试验报告	中等	无
	CT 识别	帮助从 ClinicalTrials.gov（美国临床试验注册中心）获取的记录中找到随机和半随机试验	基础	无
	ICTRP 识别	帮助从 WHO ICTRP 国际临床试验注册平台获取的记录中找到随机和半随机试验	基础	无
扩展任务	Screen4Me	帮助特定主题的 Cochrane 评价识别随机对照试验	不确定	完成 100 条 RCT 识别记录
	COVID Quest	帮助查找和描述关于 COVID-19 的研究	较难	无
	COVID Quest Lite	帮助查找关于 COVID-19 的研究	中等	无
试点任务	PICO 提取	帮助描述随机试验的人群、干预措施、对照组和结果	较难	无
	DTA 识别	帮助查找诊断测试准确性研究	较难	完成 100 条 RCT 识别记录
	表获取	绘制研究论文的关键表格	基础	完成 100 条 RCT 识别记录

扩展任务包括 Screen4Me、COVID Quest 和 COVID Quest Lite。Cochrane 为应对新型冠状病毒（COVID-19）大流行、方便研究人员和临床医生找到相关研究，创建了重要的开放获取资源 Cochrane COVID-19 研究登记册[29]，因此发起了 COVID Quest 和 COVID Quest Lite 任务，旨在通过志愿者快速识别与 COVID-19 相关的研究并为其分配一些标签。

试点任务包括 PICO 提取、DTA 识别和表获取。试点任务的目的有两种：一种是希望成为 Cochrane 流程的核心部分；另一种是进行可行性测试。

2.4 Cochrane Crowd 平台

众包平台一般具有以下几个特征：具有与人群交互的功能、具有与众包者交互的功能、具备任务相关设施、具备平台相关设施。任务相关设施是指众包平台提供的关于众包任务的设施，平台相关设施是众包平台提供的关于众包平台本身的设施[26]。Cochrane Crowd 平台除了拥有常见的众包平台功能外，在任务相关设施和平台相关设施方面具有自己的特色。

2.4.1 任务相关设施

该平台提供了“人群一致性算法”机制来聚合众包任务结果。具体为：每条记录需要 4 个连续且相同的分类才能被最终决策，当连续链被打破或出现不确定选择时，需要由经验丰富的解析者进行最后决策。另外，该平台上还使用了高亮显示功能标记标题或摘要的关键部分，辅助筛选者作出分类决定：红色高亮显示用于标记可能出现在不太相关记录中的单词，黄色高亮显示用于标记可能出现在相关研究中的关键词，蓝色高亮显示用于标记摘要中的方法部分。

2.4.2 平台相关设施

Cochrane Crowd 的特色平台相关设施为基于机器学习的 RCT 分类器。Cochrane RCT 分类器使用 Cochrane Crowd 人群数据集进行训练，用于从大型记录集中去除“噪声”，它能够提供记录为 RCT 的概率值，据此可以删除非常明显的非 RCT 记录，从而可以减少志愿者工作量并提高研究识别效率。相关研究显示，RCT 分类器能够排除 60%~80%不相关记录，同时保持超过 99%的召回率[30]。图 3 展示了 RCT 分类器在证据识别中的作用：灰色区域代表从 PubMed/MEDLINE、Embase.com、CINAHL、ClinicalTrials.gov 和 WHO ICTRP 等数据库检索出的记录，蓝色区域代表经过分类器识别后获得的潜在 RCT 记录，同时也是将要被众包筛选的部分，绿色区域代表经过人群筛选后被确认为 RCT 的记录，在分类器识别过程中可能会遗漏少量合格 RCT，但比例非常低、在可接受范围内。RCT 分类器现已构成了证据管道的一部分，部署在 Cochrane 的集成工作流中。

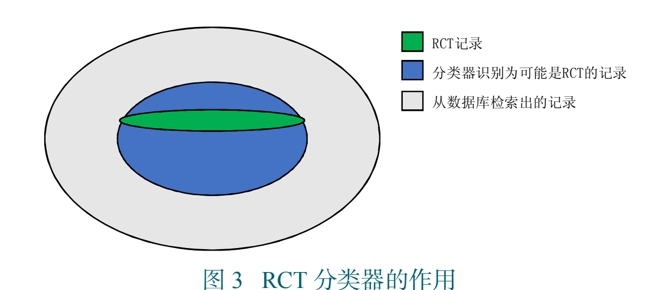


图 3 RCT 分类器的作用

Fig.3 The role of RCT classifier

2.5 质量评估

Cochrane Crowd 成功组织了一个在线社区，吸引了近 3 万名志愿者无偿帮助筛选文献，已完成了数量可观的研究识别任务，并具备可持续发展能力。一些评估显示 Cochrane Crowd 中任务识别的准确性非常高，人群敏感性为 99.1%，特异性为 99%，需要解析者识别的记录不到 20%[18]。虽然其主流任务是帮助识别特定类型研究设计的文献（RCT），但随后几项研究还评估了让志愿者根据纳入标准对特定主题系统评价所需的相关文献进行识别[19-21]，均具有良好的人群敏感性和特异性，尽管个体志愿者执行每篇摘要分类的时间要多于专家，但群体的并行工作机制足以抵消个体效率，因此人群完成任务花费的时间也远远低于作者团队。

3 Cochrane Crowd 中的众包应用对证据合成众包的启示

本文从众包者、人群、众包任务、众包平台、质

量评估 5 个维度对众包在 Cochrane Crowd 公民科学项目中的应用机制进行了分析。虽然目前对在证据合成中应用众包的研究仍处于起步阶段，但一些试验研究以及 Cochrane Crowd 的成功运行证明了在提高证据合成效率的方法中众包是一种有潜力的模式。Cochrane Crowd 中的众包机制在以下几个方面为在证据合成中使用众包带来启示。

3.1 质量控制机制

Cochrane Crowd 以任务为中心，设计了丰富的学习活动、多样化的任务访问方式、交互式定制培训模块和反馈机制，提高了志愿者正确执行任务的可能性；同时在平台层面提供一致性算法聚合人群分类结果，进一步提高了记录被正确分类的可能性。在证据合成中应用众包时，众包者必须制定适当的流程、为参与者提供全面的培训以确保资格筛选、数据提取等众包活动符合所需的高质量标准。

3.2 平台易用性

Cochrane Crowd 具有良好的交互功能、完善的任务与平台相关设施。在考察系统、网站或平台的友好性设置时，通常会使用易用性这一衡量交互式系统的重要指标。从设计角度来看，Cochrane Crowd 平台注册简单、用户界面友好、导航面板易于操作，总体上遵循了简单易用的原则，从而能够减轻志愿者的参与负担。另外平台使用了突出显示功能，能够将志愿者的注意力引导到关键短语或单词上，帮助他们作出分类决定。目前在循证科学领域，自主开发的众包平台除了 Cochrane Crowd，还有东安大略儿童医院（CHEO）研究所开发的 InsightScope[9]。这两个平台目前只能由平台开发团队以及特定人员发布任务，适用范围非常有限，亟待开发通用的证据合成众包平台，能够被各个领域需要使用众包的证据合成作者团队使用。

3.3 人机协作方式

Cochrane Crowd 通过人群的共同努力以前所未有的规模生成了大型、高质量数据集，为机器学习分类器提供了训练素材。随着人群筛选更多记录，RCT 分类器可以变得更加准确；随着分类器变得更加准确，它能够更高效地删除明显不是 RCT 的记录，从而使人群专注于需要人类智能的分类任务上，这是人机协作、相辅相成的很好示例。机器学习分类器通常需要从由专业人员生成的黄金标准分类中学习，在证据合成中使用这种分类器的最大缺点是缺乏训练数据，而众包可以作为生成这类数据的潜在方式。

4 结　语

众包是提高证据合成效率、缩短制作周期的有效方法。通过全面的参与者培训、适当的质量控制机制，可能输出高质量的众包结果并满足证据合成“黄金标准”。为了激励用户参与并促进其持续参与，应该为参与者提供明确的目标、清晰的任务和及时的反馈或奖励。

在证据合成中引入众包的兴趣和活动正在迅速增长，随着不同学科的研究人员在证据合成项目中使用众包，促进众包的新工具和平台也需要被进一步开发。未来应该对在不同领域证据合成中应用众包以及在证据合成的不同阶段使用众包展开进一步研究。

参考文献：

[1] LITTELL J H. Conceptual and practical classification of research reviews and other evidence synthesis products[J]. Campbell systematic reviews, 2018, 14(1): 1–21.

[2] DONNELLY C A, BOYD I, CAMPBELL P, et al. Four principles to make evidence synthesis more useful for policy[Z]. Nature Publishing Group, 2018.

[3] BORNMANN L, MUTZ R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references[J]. Journal of the association for information science and technology, 2015, 66(11): 2215–2222.

[4] HADDAWAY N R, WESTGATE M J. Predicting the time needed for environmental systematic reviews and systematic maps[J]. Conservation biology, 2019, 33(2): 434–443.

[5] CHALMERS I, BRACKEN M B, DJULBEGOVIC B, et al. How to increase value and reduce waste when research priorities are set[J]. The lancet, 2014, 383(9912): 156-165.

[6] TSAFNAT G, GLASZIOU P, CHOONG M K, et al. Systematic review automation technologies[J]. Systematic reviews, 2014, 3(1): 1-15.

[7] BROWN A W, ALLISON D B. Using crowdsourcing to evaluate published scientific literature: Methods and example[J]. PLoS one, 2014, 9(7): e100647.

[8] SUN Y, CHENG P, WANG S, et al. Crowdsourcing information extraction for biomedical systematic reviews[J]. arXiv preprint arXiv: 1609.01017, 2016.

[9] NAMA N, ILIRIANI K, XIA M Y, et al. A pilot validation study of crowdsourcing systematic reviews: Update of a searchable database of pediatric clinical trials of high-dose vitamin D[J]. Translational pediatrics, 2017, 6(1): 18.

[10] NAMA N, SAMPSON M, BARROWMAN N, et al. Crowdsourcing the citation screening process for systematic reviews: Validation study[J]. Journal of medical Internet research, 2019, 21(4): e12953.

[11] MORTENSEN M L, ADAM G P, TRIKALINOS T A, et al. An exploration of crowdsourcing citation screening for systematic reviews[J]. Research synthesis methods, 2017, 8(3): 366-386.

[12] PIANTA M J, MAKRAI E, VERSPOOR K M, et al. Crowdsourcing critical appraisal of research evidence (CrowdCARE) was found to be a valid approach to assessing clinical research quality[J]. Journal of clinical epidemiology, 2018, 104: 8-14.

[13] BUJOLD M, GRANIKOV V, SHERIF R E, et al. Crowdsourcing a mixed systematic review on a complex topic and a heterogeneous population: Lessons learned[J]. Education for information, 2018, 34(4): 293-300.

[14] Cochrane. About us[EB/OL]. [2022-09-04]. https://www.cochrane.org/about-us.

[15] Cochrane Collaboration. Cochrane Crowd [EB/OL].[2022-06-18]. https://community.cochrane.org/help/tools-and-software/cochrane-crowd.

[16] 尚宏利, 张思洁, 魏志鹏, 等. 循证数字人文证据整合的基本框架与具体流程研究[J]. 农业图书情报学报, 2022, 34(11): 38-47.
SHANG H L, ZHANG S J, WEI Z P, et.al. Evidence integration framework of evidence-based digital humanities[J]. Journal of library and information science in agriculture, 2022, 34(11): 38-47.

[17] PHAM B, BAGHERI E, RIOS P, et al. Improving the conduct of systematic reviews: A process mining perspective [J]. Journal of clinical epidemiology, 2018, 103: 101-111.

[18] NOEL-STORR A, DOOLEY G, ELLIOTT J, et al. An evaluation of Cochrane Crowd found that crowdsourcing produced accurate results in identifying randomized trials[J]. Journal of clinical epidemiology, 2021, 133: 130-139.

[19] NOEL-STORR A, DOOLEY G, AFFENGRUBER L, et al. Citation screening using crowdsourcing and machine learning produced accurate results: Evaluation of Cochrane's modified Screen4Me service[J]. Journal of clinical epidemiology, 2021, 130: 23-31.

[20] NOEL-STORR A, GARTLEHNER G, DOOLEY G, et al. Crowdsourcing the identification of studies for COVID-19-related Cochrane Rapid Reviews[J]. Research synthesis methods, 2022, 13(5): 585-594.

[21] NOEL-STORR A H, REDMOND P, LAMé G, et al. Crowdsourcing citation-screening in a mixed-studies systematic review: A feasibility study[J]. BMC medical research methodology, 2021, 21(1): 1-10.

[22] THOMAS J, NOEL-STORR A, MARSHALL I, et al. Living systematic reviews: 2. Combining human and machine effort[J]. Journal of clinical epidemiology, 2017, 91: 31-37.

[23] NAKATSU R T, GROSSMAN E B, IACOVOU C L. A taxonomy of crowdsourcing based on task complexity[J]. Journal information science, 2014, 40(6): 823-834.

[24] PRPIĆ J, SHUKLA P P, KIETZMANN J H, et al. How to work a crowd: Developing crowd capital through crowdsourcing[J]. Business horizons, 2015, 58(1): 77-85.

[25] ESTELLÉS-AROLAS E, GONZÁLEZ-LADRÓN-DE-GUEVARA F. Towards an integrated crowdsourcing definition[J]. Journal information science, 2012, 38(2): 189-200.

[26] HOSSEINI M, PHALP K, TAYLOR J, et al. The four Pillars of crowdsourcing: A reference model[C]//2014 IEEE Eighth International Conference on Research Challenges in Information Science (RCIS). Piscataway, NJ, USA: IEEE, 2014: 1-12.

[27] ZHAO Y X, ZHU Q H. Evaluation on crowdsourcing research: Current status and future direction[J]. Information systems frontiers, 2014, 16(3): 417-434.

[28] PEDERSEN J, KOCSIS D, TRIPATHI A, et al. Conceptual foundations of crowdsourcing: A review of IS research[C]//2013 46th Hawaii International Conference on System Sciences. Piscataway, NJ, USA: IEEE, 2013: 579-588.

[29] The Cochrane Library. Cochrane COVID-19 register of studies[EB/OL]. [2022-09-05]. https://covid-19.cochrane.org/.

[30] WALLACE B C, NOEL-STORR A, MARSHALL I J, et al. Identifying reports of randomized controlled trials (RCTs) via a hybrid machine learning and crowdsourcing approach[J]. Journal of the American medical informatics association, 2017, 24(6): 1165-1168.

作者贡献说明：

李晓：提出研究思路，研究框架设计，论文撰写；曲建升：负责论文修订；寇蕾蕾：负责论文修订。

Applications of Crowdsourcing in Evidence Synthesis: A Case Study of Cochrane Crowd

LI Xiao[1,2], QU Jiansheng[1,2,3*], KOU Leilei[4]

(1. Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000; 2. Department of Library Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100049; 3. Chengdu Library and Information Center, Chinese Academy of Sciences, Chengdu 610041; 4. Qinghai-Tibet Plateau Human Environment Research Center, Lanzhou University, Lanzhou 730000)

Abstract: [Purpose/Significance] Evidence-informed decision-making is a means to bridge the gap between research and policy and evidence synthesis has become an important tool for evidence-based decision-making in many fields. However, evidence synthesis is resource-intensive, especially when it comes to scientific knowledge on complex issues. The efficiency of evidence synthesis currently cannot meet the needs of decision makers. Crowdsourcing is seen as a potential way to improve the productivity of evidence synthesis. At present, the research and practice on the applications of crowdsourcing in evidence synthesis is still in its infancy. This study takes the application of crowdsourcing in the Cochrane Crowd citizen science project as an example to summarize the practical applications of crowdsourcing in evidence synthesis. The comprehensive analysis of the application mechanism of crowdsourcing in Cochrane Crowd project will provide certain reference and inspiration for the use of crowdsourcing in evidence synthesis, so as to improve the production efficiency of evidence synthesis and provide timely and powerful scientific information for evidence-based decision-making. [Method/Process] The application mechanism of crowdsourcing in the Cochrane Crowd citizen science project was analyzed from five dimensions: crowdsourcer, volunteers, crowdsourcing task, Cochrane Crowd platform and effectiveness evaluation, using literature research, network investigation, case analysis and other methods. Cochrane Crowd provides an easy-to-use interface for contributors to engage volunteers to participate and design , in addition to task-focused learning activities, diverse ways of accessing tasks, interactive

online training modules and feedback mechanisms to improve the likelihood of volunteers' performing tasks correctly. At the same time, an agreement algorithm is provided at the platform level to aggregate the crowd classification results, which further improves the possibility of correct classification of records. In addition, the platform has used the records identified by the crowd to build a machine-learning model called as RCT classifier which can predict how likely a new citation is to be described an RCT to reduce the manual burden. [Results/Conclusions] Crowdsourcing is an effective method to improve the efficiency of evidence synthesis and shorten the production cycle. With comprehensive participant training and appropriate quality control mechanisms, it is possible to produce high quality crowdsourcing results that meet the "gold standard" of evidence synthesis. In order to motivate volunteers to participate and promote continued engagement, participants are suggested to be provided with clear goals, clear tasks, and timely feedback or rewards. Interest and activity in introducing crowdsourcing into evidence synthesis is growing rapidly, and new tools and platforms to facilitate crowdsourcing also need to be further developed as researchers from different disciplines use crowdsourcing in the evidence synthesis projects. In the future, the application of crowdsourcing in evidence synthesis in different fields and in different stages of evidence synthesis should be further studied.

Keywords: evidence synthesis; crowdsourcing; Cochrane Crowd; evidence-based research